Protein Engineering vol.12 no.2 pp.129–139, 1999

Molecular dynamics simulation of α-lactalbumin and calcium binding c-type lysozyme

Lakshmanan K.Iyer and Pradman K.Qasba[1]

Structural Glycobiology Section, Laboratory of Experimental and Computational Biology, Division of Biological Sciences, National Cancer Institute, National Institutes of Health, Frederick, MD 21702, USA

[1]To whom correspondence should be addressed

α-Lactalbumins (LAs) and c-type lysozymes (LYZs) are two classes of proteins which have a 35–40% sequence homology and share a common three dimensional fold but perform different functions. Lysozymes bind and cleave the glycosidic bond linkage in sugars, where as, α-lactalbumin does not bind sugar but participates in the synthesis of lactose. α-Lactalbumin is a metallo-protein and binds calcium, where as, only a few of the LYZs bind calcium. These proteins consist of two domains, an α-helical and a β-strand domain, separated by a cleft. Calcium is bound at a loop situated at the bottom of the cleft and is important for the structural integrity of the protein. Calcium is an ubiquitous intracellular signal in higher eukaryotes and structural changes induced on calcium binding have been observed in a number of proteins. In the present study, molecular dynamics simulations of equine LYZ and human LA, with and without calcium, were carried out. We detail the differences in the dynamics of equine LYZ and human LA, and discuss it in the light of experimental data already available and relate it to the behavior of the functionally important regions of both the proteins. These simulations bring out the role of calcium in the conformation and dynamics of these metallo-proteins. In the calcium bound LA, the region of the protein around the calcium binding site is not only frozen but the atomic fluctuations are found to increase away from the binding site and peak at the exposed sites of the protein. This channeling of fluctuations away from the metal binding site could serve as a general mechanism by which the effect of metal binding at a site is transduced to other parts of the protein and could play a key role in protein–ligand and/or protein–protein interaction.

Keywords: calcium-binding proteins/lactose synthase/c-type lysozyme/atomic fluctuations/functionally important motion

Introduction

Two classes of proteins, α-lactalbumins (LAs) and c-type lysozymes (LYZs) have a 35–40% sequence homology and share a common three dimensional (3D) fold but perform different functions (Brew and Grobler, 1993; Sugai and Ikeguchi, 1994; McKenzie, 1996; Qasba and Kumar, 1997). These proteins are made up of two domains which are separated by a cleft. In human α-LA, for example, the residues 1–38 and 83–123 form the α-domain which predominantly consists of the α-helices, A, B, C and D (Figure 1 and Table I). The residues 39–82 constitute the β-domain which consists of β-strands, S1, S2 and S3, and the extended loop region. The positions of the four disulfide bonds, 6–120, 28–111, 61–77 and 73–91 in human α-LA, and 6–127, 30–115, 65–80 and 76–94 in equine LYZ are conserved. In addition, α-LA is a metallo-protein that has a primary calcium binding site which is required for its structural integrity, whereas only a few of the c-type lysozymes, namely, canine, echidna, pigeon and equine lysozyme bind calcium at the corresponding site (Rodriguez *et al*., 1985; Teahan *et al*., 1991; Tsuge *et al*., 1992; Grobler 1994a). This calcium binding loop is situated between the 3_{10}-helix (composed of residues 77–80 of human α-LA and the corresponding residues 80–83 of equine LYZ) and helix C. The calcium is bound in a loop at the bottom of the cleft in the α- and β-domain interface. This is different from the calcium binding EF-hand loop found in many other proteins (Stuart *et al*., 1986).

In addition to the primary calcium binding site mentioned above, recent crystallographic analysis has identified a secondary calcium binding site in human LA. The binding at this site takes place only at high calcium concentration and it does not play any structural role (Chandra *et al.*, 1998). In addition, LA has also been shown to bind other metal ions, for example, zinc but these do not play any structural role in it (Ren *et al.*, 1993). In α-LA, the calcium at the primary site is coordinated by the two backbone carbonyl oxygens, that of K79 and D84, three carboxylate side chains, that of D82, D87, D88, and two water molecules. Corresponding residues exist in the calcium binding lysozymes that coordinate the calcium. Calcium binding plays an important role in maintaining the structural integrity of human α-LA and equine LYZ (Haezebrouck *et al.*, 1992; Vandeheeren *et al.*, 1996; Wu *et al.*, 1996; Anderson *et al.*, 1997; Kuhlman *et al.*, 1997; Griko *et al.*, 1995; Hendrix *et al.*, 1996). It has been shown that the binding of calcium induces a structural transition in both α-LA and LYZ, but the exact nature of this transition has been difficult to observe experimentally. Calcium is an ubiquitous intracellular signal in higher eukaryotes and calcium induced structural transitions have been observed in a number of other proteins affecting their function (Andersson *et al.*, 1997; Laberge *et al.*, 1997; Smith *et al.*, 1996; Spyracopoulos *et al.*, 1997; Nelson and Chazin, 1998).

Despite the structural similarities, α-LA and LYZ serve very different functions. Lysozyme binds oligosaccharides and cleaves the glycosidic linkage. α-LA does not bind sugar but in the presence of sugar or sugar nucleotide, it interacts with a glycosyltransferase, UDP-galactose *N*-acetylglucosamine β-1,4-galactosyltransferase (GalT), to form the lactose synthase complex, and thereby modulates its enzymatic activity under physiological conditions (Khatra *et al.*, 1974; Bell *et al.*, 1976). A great deal is known about the structure of hen-egg white lysozyme (HEWL) and its interaction with sugars. Crystal structures of HEWL–sugar complexes have identified the sugar binding site in LYZ (for a review see Strynadka and James, 1996). A number of molecular dynamics (MD) simulations have been carried out on HEWL and its complex with

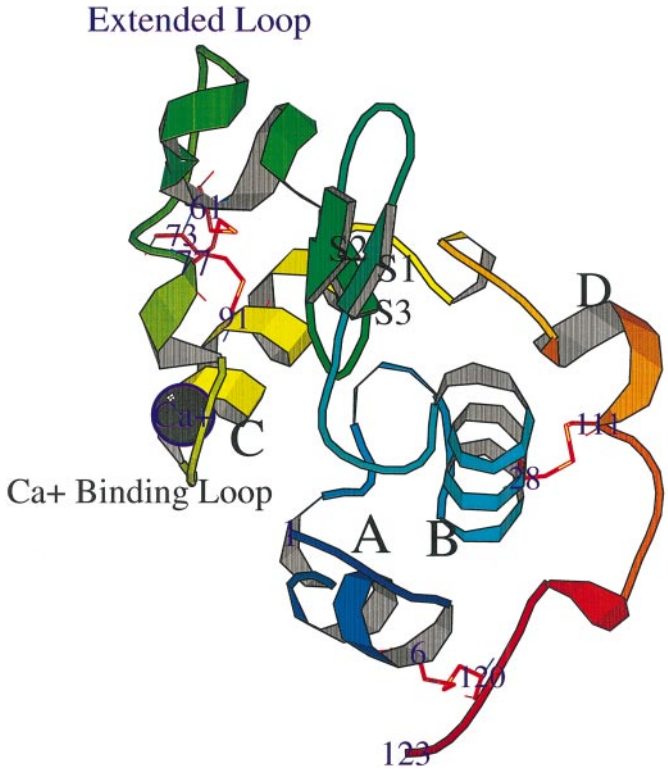

Fig. 1. The HEWL fold present in α-lactalbumin. The secondary structures, helices A, B, C and D, β-sheets S1, S2 and S3, the disulfide bridges between residues 6–120, 28–111, 61–77 and 73–91, and the Ca^{2+}-binding loop are indicated. The picture was prepared using the program MOLSCRIPT (Kraulis, 1991).

Table I. The secondary structure of human α-lactalbumin and equine lysozyme assigned using the program DSSP

Secondary structure name	Human α-LA	Equine LYZ
Helix A	K5–L11	K5–K13
Helix B	L23–S34	L25–S36
Helix C	T86–K98	D89– K98
Helix D	A106–L110	K109–H114
Strand S1	I41–E43	F43–G45
Strand S2	T48–Y50	S52–Y54
Strand S3	I55– S56	L59–S60

hexasaccharide that have yielded very interesting details regarding the role of dynamics of protein on its function (Post *et al.*, 1986, 1989; Hunenberger *et al.*, 1995; Smith *et al.*, 1995). In the case of α-LA, the NMR and X-ray structural information along with the other experimental data has helped in identifying the residues that may be important for its structure and function. The specific residues and regions of α-LA believed to be involved in its interaction with GalT have been identified (Grobler *et al.*, 1994b; Malinovskii *et al.*, 1996). The residues, F31, H32, L110 and W118 of α-LA which form an aromatic cluster (AC1) have been shown to be involved in the functioning of the lactose synthase complex (Grobler *et al.*, 1994b; numbering according to human LA). The C-terminal of α-LA from different species has been known to adopt different conformations (Acharya *et al.*, 1989, 1991; Pike *et al.*, 1996). Under varying crystallization conditions, the C-terminal residues (105–110) of human α-LA have also been shown to adopt different conformations (Harata and Muraki, 1992). This flexibility has been postulated to be important for its function. The functional role of the aspartic acid residues at the calcium binding loop has been elucidated by the construction of Asp→Ala mutants (Anderson *et al.*, 1997). The studies mentioned above constitute a wealth of data regarding the role of different residues and regions of α-LA in its structure and function.

The differences in the function of α-LA and LYZ must be attributable to the differences in their amino acid sequence, static structure and dynamic behavior. A number of reviews have compiled the results of the studies on the similarities and differences between α-LA and LYZ, in sequence, structure and function (Brew and Grobler, 1993; Sugai and Ikeguchi, 1994; McKenzie, 1996; Qasba and Kumar, 1997). Molecular dynamics simulation provides a powerful tool to understand the dynamics of a protein at atomic detail which can lead to significant insights into the atomic motions and the machinery underlying the protein function (Harvey and McCammon, 1987). The effect that the presence of calcium has on the structure of both LA and LYZ can be studied at atomic level by carrying out MD simulations of the protein in the presence and absence of calcium. In this paper, we address the issue of differences in the atomic motions of the above mentioned functionally important regions of calcium binding lysozymes and α-LAs. Towards this goal, we report the results of the simulations of the native, calcium bound, human α-LA and equine lysozyme, and discuss it in light of the existing experimental data. The binding of calcium is the step in which the two domains of the proteins get locked in place to yield the native structure. We have also conducted simulations of these proteins in the absence of calcium to understand the role of calcium in its structure and dynamics and the nature of changes that it brings about on binding to the protein. In this study, we are only concerned with the primary calcium binding site because the secondary calcium site and binding of other metal ions do not play any structural role in LA. The present simulations within their limitations (discussed in the last section) brings out the general nature of the effect that the binding of calcium has on a native or near-native structure.

Materials and methods

The crystal structures of human α-lactalbumin and equine lysozyme were the starting point for the respective simulations (Acharya *et al.*, 1991; Tsuge *et al.*, 1992). The calcium atom was not located in the crystal structure of equine lysozyme and was modeled on the basis of coordination observed in human LA. The following four systems were simulated in the present study. Human α-LA with calcium, HMLCA; human α-LA without calcium, HMLNOCA; equine lysozyme with calcium, EQLCA and equine lysozyme without calcium, EQLNOCA. Each simulated system consisted of the protein soaked in a 6 Å shell of water to simulate the water shell around the protein. All modeling and simulations were conducted using the InsightII/Discover package (Biosym Technologies Inc., San Diego, CA).

The cvff force field was used in all the simulations. Calcium was modeled by the potential type 'ca+' as defined for calcium ion in the cvff force field. The calcium coordination was modeled with distance constraints to the coordinating oxygen atoms of the protein so as to maintain the coordination. To account for the electrostatic interaction, a charge of +2 was assigned to the calcium atom. The following protocol was used in the simulation. In the initial 500 steps of energy minim-

Table II. Average values for the r.m.s.d./res for all atom (1st row) and the backbone atoms N,Cα,C only (2nd row) in all the four systems with their standard deviations in parentheses

EQLCA	EQLNOCA	HMLCA	HMLNOCA
2.18 (0.6)	1.33 (0.6)	1.7 (0.42)	0.91 (0.27)
2.00 (0.57)	1.08 (0.58)	1.48 (0.33)	0.92 (0.21)

ization, all the heavy atoms, of both the protein and water molecules, were held fixed and the hydrogens allowed to move so as to optimize their spatial position. Following this, only the heavy atoms of the protein were held fixed and the system was energy minimized for another 500 steps, essentially to reorient the water molecules around the protein. Finally, the whole system was minimized to arrive at a starting structure for the simulation. Subsequently, the system was equilibrated at 300 K for 50 picoseconds (ps) during which time the potential energy of the system reached a stable value. Following this a 250 ps production was carried out (for HMLCA this period was 243 ps). The other parameters used in the simulation were as follows. A cut-off of 15 Å was used to evaluate the non-bonded interactions. A time step of 1 femtosecond (fs) was used to integrate the equation of motion. The calcium coordination constraints were maintained during the whole simulation period. The trajectory data was saved at every 0.25 ps for analysis. The root mean square deviation (r.m.s.d.) of each residue during the simulation period was calculated. In the case of backbone atoms, only N, Cα and C atoms were considered since the carbonyl oxygen movements are similar to the side chain (Karplus and Post, 1995). Snapshots of the protein at every 25 ps were extracted. These structures were analyzed using the program DSSP (Kabsch and Sander, 1983). Visual inspection of these superimposed structures were also done in order to study the motions of the protein in detail and the results are discussed in the following sections.

Results

Analysis of average r.m.s.d./residue

The average r.m.s.d. of a residue in the protein for the entire course of the simulation is presented in Table II. In these entries, two averages have been calculated; one over the time period of simulation and the other over the number of residues in each protein. On average, a residue in equine LYZ shows more fluctuation than in human α-LA, in both the calcium bound and free forms (Table II). This is also evident in Figure 2 which shows the average r.m.s.d. of each residue for the entire period of simulation. All the graphs in Figure 2 show a very similar profile which is characteristic of the c-type lysozyme fold, and which has been seen in the B-factors of the crystal structures, as well as in the MD-simulation results on HEWL (Strynadka *et al.*, 1991; Song *et al.*, 1994; Karplus and Post, 1995; Hunenberger *et al.*, 1995). Except for the crystal structures of α-LA, none of the other X-ray structures or molecular systems simulated had calcium bound to the protein in these studies (the calcium was not located in equine lysozyme crystal structure). In Figure 3, we compare the r.m.s.d./res of the backbone atoms to the already published, experimentally determined amide protection factors for both human α-LA and equine LYZ in their native states (Schulman *et al.*, 1995; Morozova-Roche *et al.*, 1997). The backbone r.m.s.d. values correlate well with the observed protection factors of both the proteins (see Discussion for details). This validates the protocol used in the simulation, especially the modeling of the calcium in the native state of both the proteins. The results of the present simulation presented in Table II show that on average, the fluctuations in both the backbone, as well as for the whole residue, are more in the calcium bound, halo-protein (EQLCA, HMLCA) than in the apo-proteins (EQLNOCA, HMLNOCA). The usage of the term halo- or apo-protein in the context of the present simulation implies the presence (native) or absence of calcium in the simulated system, respectively. The power of simulations to remove calcium in an artificial, yet simple way, has been exploited in these simulations to understand the effect of calcium on the structure and dynamics of the protein. Hence, the calcium deficient systems simulated here do not necessarily correspond to the experimentally observed apo-form of the proteins. In general, the side chains in a protein are more flexible than the main chain, and a comparison of the top and bottom panels of Figure 2 clearly indicates that most of the motions in both the proteins are contributed by the side chain. The bottom panel of Figure 2 also shows that in the absence of calcium, the fluctuations in the backbone of LYZ and α-LA are very similar. In contrast, the presence of calcium in lysozyme increases the magnitude of motion in the protein as compared with lactalbumin, especially around the N- and C-termini. Due to the close proximity, the N-terminus of both the proteins can directly interact with the calcium binding loop. The C-terminus of both the proteins in turn lie closer to the N-terminus and is coupled to it via the disulfide bond, 6–120 in human α-LA and 6–127 in equine LYZ. Hence, the influence of calcium on the motion of the N-terminus is also propagated to the C-terminal region. However, the magnitude of fluctuation is smaller for α-LA as compared with LYZ. The regions of LYZ and α-LA that show lower value and very similar behavior in the r.m.s.d. plot correspond mostly to the secondary structure elements in the protein as shown in Figure 1 and marked in Figure 2.

In the calcium bound form, the peak in r.m.s.d./residue for both the proteins are in the loop connecting helices A and B, the exposed turn connecting strands S1 and S2, the extended loop and the C-terminus of the protein (Figure 2). In the calcium bound form, helix A, the N-terminus of helix B and loop connecting helices A and B show more fluctuations in EQLCA than HMLCA. A similar result holds for the loop connecting strands S1 and S2. The helix C and the residues connecting it to helix D, which line one side of the cleft, show larger r.m.s.d. in EQLCA than in HMLCA. In EQLCA, helix D is the only region in its C-terminal end that shows lower r.m.s.d. which is comparable to that of HMLCA. Although the C-terminal region of both the proteins is flexible, in LA it seems to be less flexible than lysozyme. The behavior of the C-terminus in relation to the formation of the lactose synthase complex and the interactions with the substrates have been discussed separately. In the absence of calcium, in EQLNOCA and HMLNOCA, the significant differences in fluctuations are seen in the extended loop and in the 3_{10}-helix that is N-terminal to the calcium binding loop. The exposed turn connecting strands S1 and S2 also show some reduction in motion. The most striking result is the very close behavior of the N- and C-terminus of both the proteins (Figure 2). The δ(r.m.s.d.) plot for the backbone atoms of both the proteins summarize these results about the effect of the presence of calcium on the protein dynamics (Figure 4).

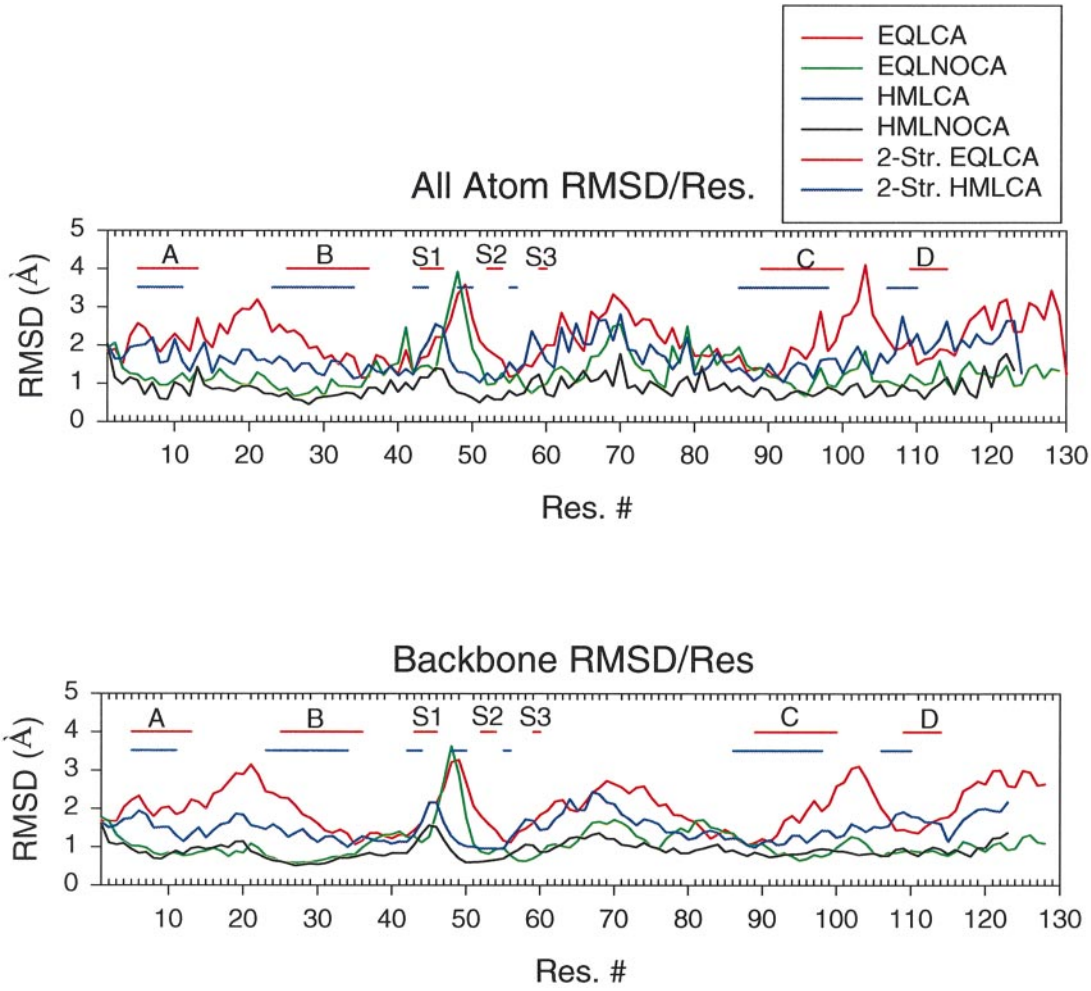

Fig. 2. The r.m.s.d./residue plot for the MD-simulated structures of human α-lactalbumin, with (HMLCA) and without calcium (HMLNOCA), and equine lysozyme, with (EQLCA) and without calcium (EQLNOCA), for all atoms of a residue and the backbone. The secondary structures, helices A, B, C and D, and β-sheets S1, S2 and S3, of both proteins are also marked.

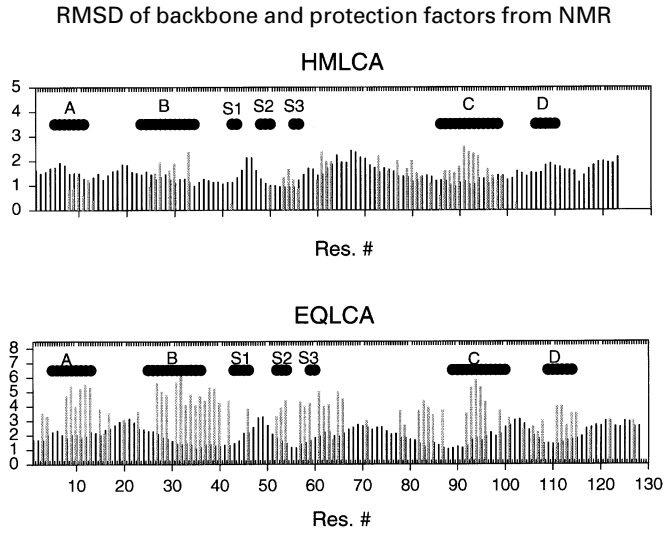

Fig. 3. Comparison of the experimentally observed amide protection rates and the r.m.s.d./res of the backbone for human α-LA (HMLCA) and equine LYZ (EQLCA). An arbitrary scale on the Y-axis has been used to plot protection factors (grey bars) and the observed r.m.s.d. (black bars). The secondary structures in the two proteins are indicated by the black horizontal bars on the top portion of each panel.

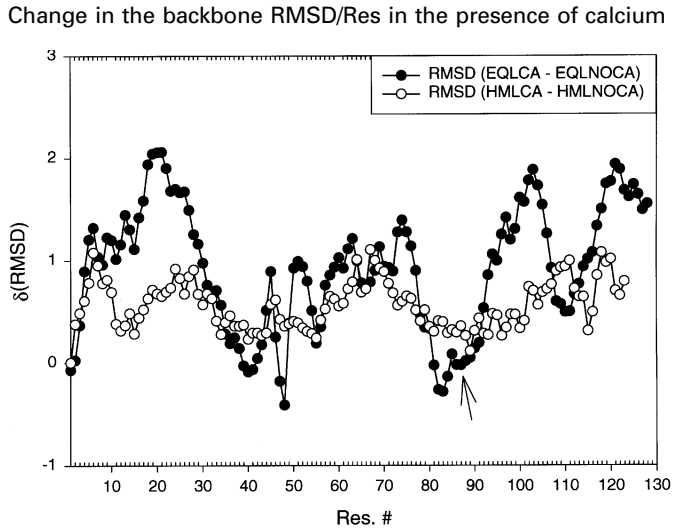

Fig. 4. Change in r.m.s.d./residue for the two proteins, human α-LA and equine LYZ, on binding calcium. The values were arrived at by subtracting the r.m.s.d./res for the calcium free simulation (EQLNOCA, HMLANOCA) from that of the calcium bound (EQLCA, HMLCA) for each of the residues in the protein. The arrow indicates the calcium binding loop.

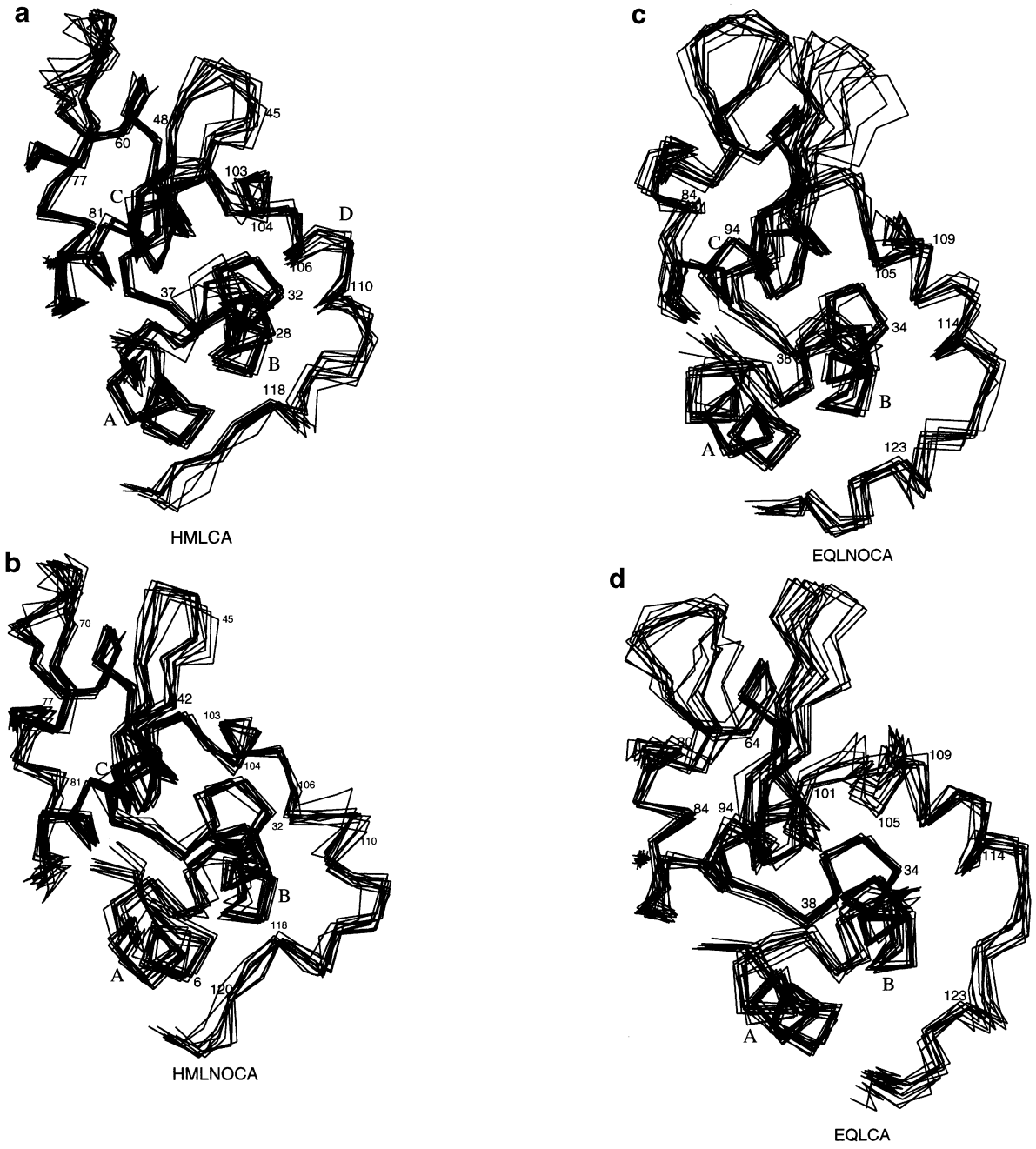

Fig. 5. Superposition of every 25 ps snapshot from the MD-simulation trajectory for the four simulated systems: (**a**) HMLCA; (**b**) HMLNOCA; (**c**) EQLCA; (**d**) EQLNOCA. The helices in the proteins are labeled. Some of the important residues that are referred to in the Discussion are also numbered.

Analysis of snapshots at every 25 ps

Secondary and tertiary structure variation

The motion at the secondary and tertiary structure levels of a protein are very intimately connected to the functioning of the protein. Snapshots taken from the dynamic trajectory at every 25 ps are superimposed and presented in Figures 5a–d. Although a large portion of the secondary structures in all the four simulated systems remain unaltered (Figure 5a–d), there are, however, significant variations in the regions adjacent to these secondary structures. The helices, A, B and C show very little unwinding in HMLCA. Helix A in HMLCA changes its orientation during the simulation period, and the loop connecting helices A and B shows fluctuations (Figure 5a). Helix B, which is the most buried of the helices, moves as a rigid unit. Its N-terminus is almost fixed. The center of the helix is also clamped by the disulfide bond, 28–111. The motion of the helix is in the C-terminus only, to which the residues F31, H32 are attached. The peaks in r.m.s.d./res between residues 45–50 (Figure 2) are due to the exposed loop between the strands S1 and S2, and those between residues 60–80 are due to exposed coiled loop, which is restrained by the disulfide bond 61–77. The 3_{10}-helix beginning at 77 and N-terminal to the calcium-binding loop adopts an α-helical conformation and remains rigid during the course of the simulation.

Helix C is fixed at the N-terminal end by the calcium-binding loop and pivoted in the middle by the 73–91 disulfide bond, so that most of its movement is in the C-terminal end. The residue Y103 positioned in the cleft at a site corresponding

to the sugar binding site A–D of LYZ, maintains its packing in the cleft and keeps that part intact. Helix D extends at least through three residues, 105–108, and moves as a rigid unit and affects the orientation of W104 at the cleft.

In HMLNOCA (Figure 5b), the helix A remains stable and moves as a rigid unit. The motion of the most buried helix, helix B, is similar to that in HMLCA. Unlike HMLCA, the stretch of residues 40–70 does not show big peaks in fluctuation as seen in Figure 2. This may be due to the following reasons. Helix C in HMLNOCA does not show as much motion as in HMLCA, and stays as a rigid unit. The most notable changes in HMLNOCA are in the following regions. The 3_{10}-helix, N-terminal to the calcium-binding loop does not adopt α-helical conformation as seen in HMLCA. The calcium-binding loop and its neighboring residues are flexible. There is dissipation of fluctuation in this region itself, so that the helix C remains more or less rigid. This also affects the sheets and the loops connecting them and they do not show as much motion as in HMLCA. The orientations of Y103 and W104 are maintained in the cleft during the simulation period of HMLNOCA. However, helix D, residues 106–110, shows structural transitions between the α-helical (mostly adopted in HMLCA) turn and strand conformations. The C-terminal end of the protein shows fluctuations similar to HMLCA.

In EQLCA (Figure 5c), helices A, B and C remain stable during the simulation period. Helix A moves during the simulation and hence there is a large r.m.s.d./Res between the residues 10–20 as shown in Figure 2. Helix B moves similar to the other three systems, clamped in the middle by the 30–115 disulfide bond. The three strands, S1, S2 and S3, move in tandem as rigid units but fluctuation is chiefly due to movement of S1 and the loop connecting S1 and S2. The coiled loop also fluctuates (disulfide bond at 65–80). The residues 80–84, which are towards the N-terminal of the calcium-binding loop, remain in the α-helical conformation for the whole simulation period. The helix C exhibits a rigid movement along its length and breaks around the disulfide bond, 94–86. The residues 93–95 adopt a turn conformation. This leads to fluctuations in the residues, 100–109, which line the inter-domain cleft. Helix D is stable, its C-terminal end fixed by the 115–30 disulfide bond, but due to the motion of its N-terminus it reorients a great deal (in concert with helix B) during the simulation. The rest of the C-terminal end of the protein, beyond residue 115, shows large fluctuations.

In EQLNOCA, the motion of helix B (Figure 5d) is similar to that of helix B in LA. The strands S1 and S2, and the loop connecting them show large fluctuations (see also Figure 2). The exposed coiled loop also shows large fluctuation. On removal of calcium, the residues 80–84 remain helical for most of the simulation period but go into the turn conformation during the last 25 ps of simulation. These residues move and reorient as a rigid unit during the course of the simulation. The N-terminal half of the helix C packs against the 3_{10}-helix preceding the calcium-binding loop. The fluctuations in the C-terminal half of helix C affects the residues lining the cleft and those in helix D. Interestingly, helix D remains helical unlike its counterpart in LA, and does not change much in length but changes in orientation and affects both the residues in the cleft and the aromatic cluster.

Thus, in all the four systems the dynamics of the secondary structures of the α- and β-domains exhibit similar behavior which reflect the common HEWL fold with the characteristic conserved disulfide bonds (Figure 1). As shown in Figures 6a and b, calcium-binding in HMLCA and EQLCA introduces rigidity to its neighborhood and the dynamic portions of the protein are the residues distant from the binding site. This results in large fluctuations of the loops connecting S1 and S2, and the C-terminus of helix C. This may be the most probable reason for the large fluctuation seen in the C-terminal portion of the protein.

As expected, the calcium coordination imposes a rigid geometry in the loop connecting the 3_{10}-helix and helix C, and also affects the interface residues. During the course of simulation the aspartic acid residues, D85, D87, D89 and D91 of EQLCA and D82, D84, D87 and D88 of HMLCA retain the side chain orientation at their initial χ1 values (data not shown). Residues D91 of EQLNOCA and D88 of HMLNOCA show changes in χ1 angle. In EQLNOCA χ1 of D91 goes to approximately the –180° region during the first half of the simulation period. However, after about 100 ps it changes its value to around to –60° that is seen in the corresponding residues in HMLCA and EQLCA. In HMLNOCA, the fluctuations of χ1 of D88 are more than those of the other aspartic residues in the calcium binding loop in all the four systems. Thus, there seems to be close interplay between the structure of the protein and the geometry and integrity of the calcium coordination. Another interesting observation is the water mediated interaction of K79 of LA with the bound calcium. In the crystal structure, the K79 side chain is oriented towards the solvent. However, in HMLCA, it comes within 6 Å of the calcium atom once during the course of the simulation, where as the minimum value of the corresponding distance in EQLCA is 7.7 Å (data not shown). The calcium coordination is well maintained during the course of the simulation.

Aromatic cluster (AC1)

In HMLCA (Figure 7a) the aromatic cluster is maintained intact during the course of the simulation (see Figure 1 and 7a). The side chains of the residues W118 and F31, which directly interact with each other, show some rigid body motion but do not show large variation in their accessible surface area (data not shown). The other two residues, H32 and Y36, do not show any significant motion. On the removal of calcium, HMLNOCA, the W118 and F31 which interact directly with each other get buried and are fixed in their positions. The residue H32, unlike that in HMLCA, shows more flexibility. This subtle change in the aromatic cluster behavior is primarily due to the difference in behavior of the flexible loop 105–110 and the residues following it at the C-terminus of LA.

The corresponding aromatic cluster in lysozyme is comprised of residues Y34, F38 and Y123 (equine LYZ numbering) (Figure 7b). In EQLCA, the E33 hydrogen bonds with Y123 and this interaction is different from the interaction between the rings of F31 and W118 of LA. In lysozymes, Y34 can also interact and possibly form an H-bond with E33. The residues E33, Y34 and F38 retain their degree of partial exposure to the solvent, so does Y123 (the H-bond interaction with E33 being the reason) and E122. In the calcium free LYZ, EQLNOCA, these residues are relatively more fixed. Although Y34 shows more motion in EQLNOCA than in EQLCA, it is, however, not a large motion. The residues E122 and Y123 are relatively more well buried in EQLNOCA than in EQLCA. The reason for the observed immobility of Y123 may be that it is required to help stabilize the disaccharide at the E and F sites of lysozyme.

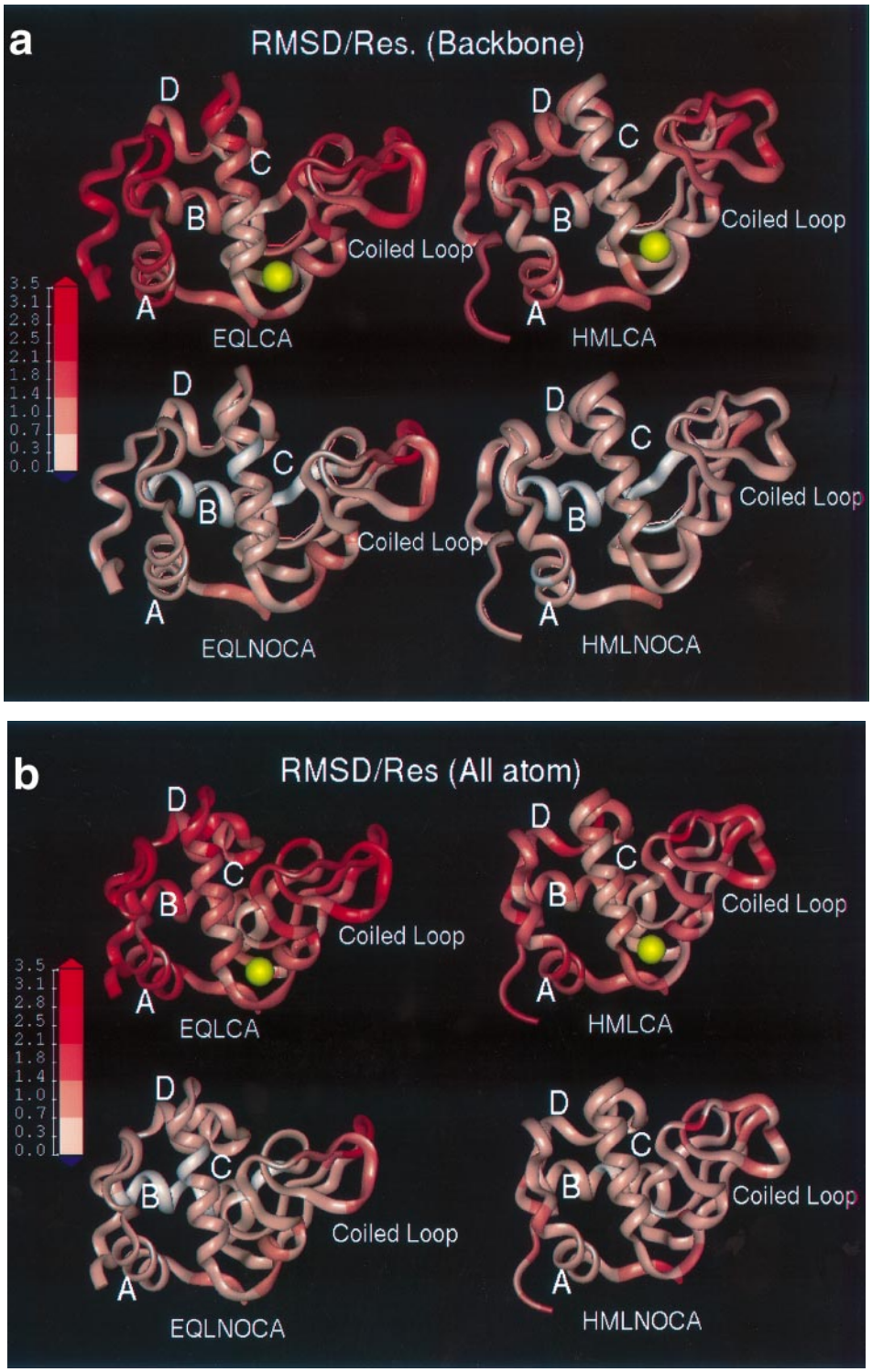

Fig. 6. Ribbon diagrams of the two proteins showing the average r.m.s.d. that a residue attains during the entire simulation period. The magnitude of the r.m.s.d. is represented by the intensity of the color in the spectrum. The scale for the r.m.s.d./res is shown by the color bar on the left. Equine lysozyme with calcium (EQLCA, upper left) and without calcium (EQLNOCA, lower left); human LA with calcium (HMLCA, upper right) and without calcium (HMLNOCA, lower right). (**a**) Spectra of backbone r.m.s.d./res; (**b**) spectra of all atom r.m.s.d./res.

Behavior of the flexible loop, 105–110, of α-lactalbumin and the C-terminal region

The flexible loop of α-lactalbumin, residues 105–110, remains in the α-helical conformation for most of the simulation period, but the residues 109 and 110 show transition into the turn region (Figures 7a and b). This is perhaps the reason for the fixed orientation of H32, the functionally important residue of aromatic cluster in HMLCA. This also helps to maintain the shape of the cleft in α-LA. The rest of the residues do not show any helical character except the snapshot at 243 ps, in which residues 116–118 show 3_{10}-helical character. Upon removal of calcium in human α-LA (HMLNOCA), helix D

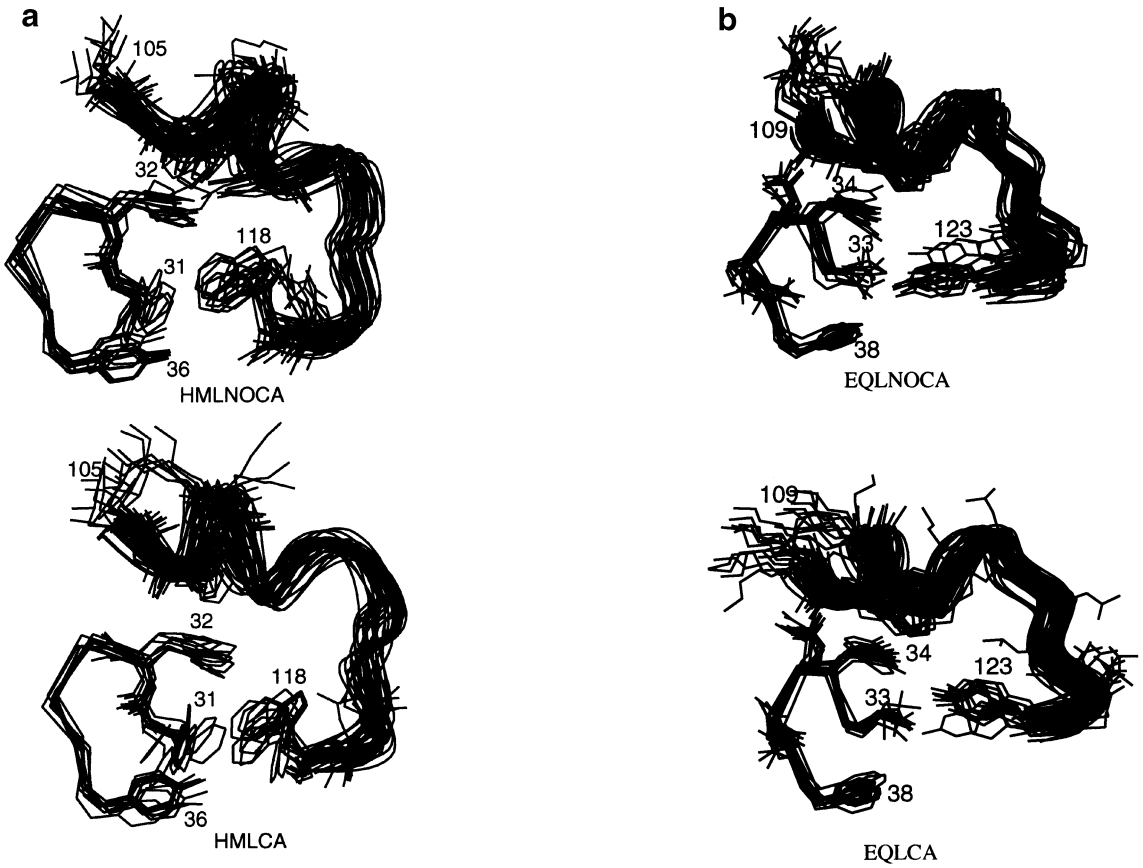

Fig. 7. (**a**) The 25 ps superimposed snapshots of the aromatic cluster, residues F31, H32, W118 and Y36, in the simulation of human α-LA with (HMLCA) and without calcium (HMLNOCA). The ribbon diagram of the backbone of C-terminal residues, 105–118, is also shown. Notice the unwinding of the single turn of helix D, starting from residue 105 in HMLNOCA as compared with HMLCA. (**b**) The 25 ps superimposed snapshots of the aromatic cluster, residues Y34, F38 and Y123, in the simulation of equine LYZ with (EQLCA) and without calcium (EQLNOCA). The ribbon diagram of the backbone of C-terminal residues, 109–123, is also shown. Notice there is no unwinding of helix D, both in EQLNOCA and EQLCA.

unwinds during minimization. During the course of the simulation, the helical character shifts down the sequence, with the residues 109–112, showing α-helical character during the course of the simulation and thus changing the burial of W118 and also the shape of the cleft corresponding to the F-site of lysozyme. The residues 116–118 show 3_{10}-helical character in a few snapshots, thus indicating that this is an intrinsic character of the amino acid sequence of α-LA. It is also interesting to note that the residues 101–103, which line the cleft, adopt 3_{10}-helical character in minimized HMLCA, but goes into S (strand) or T (turn) conformation during the simulation. In the absence of calcium, the corresponding stretch of residues show 3_{10}-helical character transiently during many of the 25 ps snapshots analyzed. This would modify the shape of the active site, cleft sites B–D.

In equine lysozyme, in the presence of calcium, the α-helix is maintained in the corresponding stretch of residues 109–112, and residue 113 and 114 show some unwinding into the turn region. Similar behavior is seen in EQLNOCA, which shows that the absence of calcium does not destabilize helix D of lysozyme. Residues 105–109, lining the active site cleft, show behavior similar to the corresponding residues of α-LA. The correspondence holds both in the presence and absence of calcium. Thus, absence of calcium seems to induce the formation of 3_{10}-helix in the cleft near sugar binding sites B–D. The rest of the C-terminal residues of both EQLNOCA and EQLCA do not show any helical character.

Discussion

Equine lysozyme and human α-lactalbumin share the same HEWL fold but serve different functions. The difference in function must be due to the amino acid sequence differences and consequently the dynamics of the respective proteins. The present simulations are the first attempts to probe the dynamics of calcium-bound α-LA or LYZ using MD simulations and correlate that to the functional aspects of the protein. The amide exchange measurements have been reported for the native and molten globular states of human α-LA and equine LYZ (Schulman *et al.*, 1995; Morozova-Roche *et al.*, 1997). A comparison of the published amide protection factors of the native protein and the backbone r.m.s.d./Res observed in the present simulations for the proteins in the presence of calcium are presented in Figure 3. The result clearly show the correlation of the region of low r.m.s.d./Res with the regions showing high amide protection factors in the NMR experiments. The graph also brings out the subtle differences in the dynamics of native equine LYZ and human α-LA. In equine lysozyme, the most protected helix is helix B, followed by helix A, C and D. This correlates fairly well with the variation in fluctuations seen for the backbones of the helices. The experimental data shows that the residues 61–63, and 78 and 79 in the extended loop, which are situated close to the disulphide bonds 65–81 and 77–95, are significantly protected. These residues also show lower r.m.s.d. in the simulations. The calcium binding ligands, D85 and D87, that showed significant

exchange protection also shows lower mobility during the course of the simulation. In human α-LA, helix C is the most protected, followed by helix B and helix A, and these are in agreement with the r.m.s.d. fluctuations seen in the simulation. In the β-domain, the regions that show higher amide protection are the residues around the disulfide bonded cysteines, 61 and 77, and 73 and 91. In the simulation, these residues also show lower fluctuations in the backbone. The dynamics clearly shows the fluctuations in helix C of equine LYZ and its affect on its C-terminus being exposed to the solvent. The excellent correlations seen between the experimental and simulation results confirms the validity of the simulation, especially the modeling of the calcium in the native state of the proteins. This also provides an atomic picture of the details of the motion taking place in these proteins.

Dynamics of the protein plays a crucial role in the binding of a ligand to the protein (McCammon and Harvey, 1987). The simulation results shown in Figure 2 show that the r.m.s.d./Res profile of motion is typical of the HEWL fold. Most of the differences in the motion of LA and LYZ are chiefly due to the insertions present in the sequence of LYZ when aligned to LA (LYZ is usually 129 residues long and LA is 123 residues long). The loop connecting helices A and B has a two residue insertion in LYZ as compared with LA. This results in the high fluctuation of this loop, the C-terminal half of helix A and N-terminal half of helix B in EQLCA as compared with HMLCA. The other significant fluctuation in EQLCA and EQLNOCA are in the exposed loop connecting strands S1 and S2. This loop is very much like a flap on the sugar binding cleft of LYZ and its motion may be important for sugar binding. Although the corresponding region in lactalbumin also shows fluctuations but they are of much lesser magnitude than in lysozyme. It is also interesting to note that this loop contains the glycosylation site in various α-LAs and some α-LAs are glycosylated (Brew and Grobler, 1993; Qasba and Kumar, 1997). The exposed coiled loop (see Figure 1) in both LA and LYZ show large fluctuations, both in the side chain as well as in the backbone (Figure 2). Interestingly, in the simulation of the HEWL–sugar complex, it was seen that substrate binding affected the corresponding regions of LYZ, though the coiled loop is not in direct contact with the substrate (Post *et al.*, 1989). A similar effect was also seen in the crystal structure of the HEWL–trisaccharide complex (Strynadka *et al.*, 1991). On removal of calcium, EQLNOCA still shows more fluctuations than HMLNOCA, indicating that the movement in this region is an intrinsic property of the LYZ sequence that is important for its function involving the binding of sugar. Another interesting observation is the larger amplitude of motion of the C-terminus of helix C and the loop connecting helices C and D in EQLCA as compared with HMLCA. Helix C occurs at the interface of the domains and the loop connecting it to helix D lines one side of the sugar binding site. The motion in the corresponding residues has also been observed in the simulations of native HEWL (Post *et al.*, 1986). However, in the simulations of hexasaccharide bound complex, these residues were found to be less mobile implying that this motion is important for lysozyme activity (Post *et al.*, 1989). The absence of the corresponding motion in LA may be due to the following reason. Y103 at this site in α-LA packs well with other residues in the cleft and is also responsible for blocking the entry of sugar in this cavity. During the course of the simulation this residue is well buried in the cleft. Thus, there are very clear differences in the dynamics of α-LA and the calcium binding c-type lysozymes which are important for their different function. Interestingly, experimental confirmation for the involvement of differences in dynamic fluctuations of homologous proteins was recently provided by the work of Zavodszky *et al*. (1998) in which the relation of the flexibility of the protein to its activity was investigated. They showed that the differences in dynamics of the thermophilic and mesophilic homologues of 3-isopropylmalate dehydrogenase (IPMDH) were crucial to the optimal temperature at which they function. The experimental results showed that the thermophilic IPMDH is significantly more rigid at room temperature than its mesophilic counterpart. However, both the enzymes showed very similar fluctuations at temperatures near their activity optima. Thus, the differences in dynamics elucidated in the current study of the homologous proteins, α-LA and calcium binding LYZ may be directly correlated with their different functions.

In a number of proteins, including α-LA and equine LYZ, calcium binding induces structural transitions (Haezebrouck *et al.*, 1992; Andersson *et al.*, 1997; Laberge *et al.*, 1997; Smith *et al*., 1997; Spyracopoulos *et al.*, 1997). For equine LYZ and α-LA a reduction in the protein hydrophobicity, as measured by binding to a hydrophobic column on calcium binding, has been observed (Lindahl and Vogel, 1984; Haezebrouck *et al.*, 1992). The environment around the hydrophobic residues W28 and W108 of equine LYZ has been shown to be altered on calcium binding (Morozova *et al.*, 1991; Tsuge *et al.*, 1991). The corresponding hydrophobic residues of α-LA are the residues Y103 and W104, which have been suggested to be exposed to the solvent in a calcium dependent way (Koga and Berliner, 1985). In the case of EQLCA, which shows large fluctuations compared with EQLNOCA, the accessible surface area (ASA) of the hydrophobic residues V99 and W108 which line the cleft, is reduced indicating its burial due to the movement in this loop as discussed above. In the 250 ps snapshot, it is 62 $Å^2$ in EQLNOCA compared with around half this value, 23 $Å^2$ in EQLCA. In the corresponding regions of α-LA, the residues 100–105, there are four hydrophobic residues, namely I101, Y103, W104 and L105, which show no such dramatic difference in the ASA; 280 $Å^2$ of HMLCA at 243 ps compared with 215 $Å^2$ of HMLNOCA at 250 ps. In fact, in this case, hydrophobic ASA of this region of the protein in the presence of calcium is more than in its absence. In comparing these kinds of results one may not expect a perfect correlation between the system simulated here and the actual experiments for the following two reasons: (i) the time scales of the simulations are very small as compared with the actual experimental time scales and (ii) as mentioned earlier, the simulated systems, EQLNOCA and HMLNOCA, in which calcium is not present, need not necessarily correspond to the apo-form of the respective proteins. However, the simulations were carried out to understand the nature of influence that calcium has on the structure and dynamics of these two proteins. Recently, MD simulations with and without calcium were reported for parvalbumin that contains two classical EF-hands which bind two calcium ions. Comparison of the end point structures of a 200 ps simulation of the calcium deficient system with the native state showed perturbation in helices that is in agreement with the experimental results (Laberge *et al.*, 1997). Thus, the protocol and time periods of simulations reported in the present study seem to be reasonable for addressing the question of the effect of calcium binding on

protein structure and dynamics. Additionally, it has also been suggested that metal binding effects the dynamics of the protein, rather than its equilibrium properties (Jernigan *et al.*, 1994).

Recently, the role of aspartic acid residues in the calcium coordination, and hence in the structure and function of LA, was investigated by site-directed mutagenesis (Anderson *et al.*, 1997). It was shown that the residues D87 and D88 were very crucial for primary calcium binding, structural stability and function of LA. As discussed in the results section, the trajectories of $\chi 1$ of D88 of LA (corresponding residue D91 of LYZ) is more affected on the removal of calcium than the other calcium coordinating residues. Thus, the geometry of the calcium binding loop seems to influence the motion of the residue at this position. Indeed, it has been proposed that D87/D88 docks the calcium ion in the initial stage of calcium binding, followed by the capping of the coordination by the other residues. It was also seen that the mutation of residue 79 from Lys to Ala also altered the structure and stability of the mutant. It has been proposed that K79, conserved in all LAs, may interact with the calcium ion through water molecules (Qasba and Kumar, 1997). Indeed, such an interaction in HMLCA is seen in our simulation, though transiently (data not shown). This may be the reason for the observed alteration in structure and stability of the K79A mutant of LA. Another interesting observation has been the nearly 100-fold increase in the absolute calcium affinity of the recombinant LA compared with that isolated from milk (Anderson *et al.*, 1997). The recombinant LA had an extra methionine residue added at the N-terminus and it was speculated that this may be interacting with the calcium binding loop. This interaction is clearly seen in the present simulations. In our simulations, we see that in the presence of calcium, the fluctuations in the N- and C-terminus of both LA and LYZ increase compared with that in its absence (Figures 2 and 3), implying a direct effect of the calcium binding on these termini. Our simulation also shows that the rigidity imposed upon by the primary calcium binding in HMLCA maintains the geometry of the adjacent secondary calcium binding site that was recently identified in the crystal structure of human LA (Chandra *et al.*, 1998).

Finally, the present investigation brings out the nature of the 'conformational change' associated with calcium binding. From Figure 6a and b, one can see that the binding of calcium increases the magnitude of fluctuations in the proteins at regions far from its site of binding (upper panels in Figure 6a and b). In the absence of calcium, the inherent fluctuations in the protein corresponding to the room temperature motion is dissipated almost uniformly over the protein (lower panels in Figure 6a and b). The introduction of calcium not only freezes the region around this site but also leads to channeling of the fluctuation to sites away from it. The fluctuation grows as it moves away from the calcium site and, as mentioned earlier, peaks at the exposed sites of the protein (Figures 2 and 4). Though the time period of simulation presented here is only 250 ps, we can clearly see the nature of changes that may occur in a protein on binding of calcium. This raises the interesting possibility of being a general mechanism by which the protein dynamics and function are modulated by metal binding. The function of the metal binding site may not only be to maintain the geometry of the cleft but also to make the protein flexible away from the binding site, and at regions of the protein that are responsible for sugar binding in the lactose synthase complex. This may be the general mechanism by which signals induced by metal binding to protein are transduced from one part of the protein to regions far away in the protein that are involved in protein–ligand and protein–protein interactions. In fact, a number of processes at the cell surface involve stimulus due to calcium or other metal binding (Maurer *et al.*, 1997) leading to alterations on the proteins at the cell surface and their interactions.

As mentioned earlier, the ability of the region of α-lactalbumin (105–110) to undertake loop and helical conformation has been seen in earlier experiments and has been hypothesized to be important for its interaction with GalT (Acharya *et al.*, 1989, 1991; Harata and Muraki, 1992; Pike *et al.*, 1996). The present simulation results show the nature of the transition that takes place, though in this case it is induced by the absence of calcium. Despite these structural transitions, no change in the orientation of the residues in the aromatic cluster, which is also important for the involvement of α-LA in lactose synthase function, is observed during the simulation period. One of the most interesting observations is the retention of α-helical conformation for residues 109–114 of equine lysozyme, even in the absence of calcium, unlike the corresponding residues, 105–109, in α-LA. This C-terminal part of LYZ, which is situated away from the sugar binding sites A–D, does not have a direct role in its function. However, it does help in keeping the aromatic cluster and residues in the E and F sites in place. This may be important for the processing of oligosaccharides by LYZ because in the crystal structure of the HEWL–hexasaccharide complex (Song *et al.*, 1994), the cleaved disaccharide was found bound to the region corresponding to E and F sites and indicates a possible site for tweaking LA so that it binds a sugar.

Acknowledgements

The authors thank Dr R.L.Jernigan for a critical reading of the manuscript and invaluable suggestions to improve it. We also thank Drs N.Pattabhiraman and Karol Miaskiewicz for fruitful discussion during the course of this work. All computations were carried out at the Frederick Biomedical Supercomputer Center (FBSC) at the National Cancer Institute–Frederick Cancer Research and Development Facility (NCI–FCRDC), Frederick, MD, USA. L.K.I. is a Fogarty Visiting Fellow at NIH.

References

Acharya,K.R., Stuart,D.I., Walker,N.P.C., Lewis,M. and Philips,D.C. (1989) *J. Mol. Biol.*, **208**, 99–127.

Acharya,K.R., Ren,J., Stuart,D.I., Philips,D.C. and Fenna,R.E. (1991) *J. Mol. Biol.*, **221**, 571–581.

Anderson,P.J., Brooks,C.L. and Berliner,L.J. (1997) *Biochemistry*, **36**, 11648–11654.

Andersson,M., Malmendel,A., Linse,S., Ivarsson,I., Forsen,S. and Svensson,L.A. (1997) *Protein Sci.*, **6**, 1139–1147.

Bell,J.E., Beyer,T.A. and Hill,R.L. (1976) *J. Biol. Chem.*, **251**, 3003–3013.

Brew,K. and Grobler J.A. (1993) *Adv. Dairy Chem.*, **1**, 191–229.

Chandra,N., Brew,K. and Acharya,K.R. (1998) *Biochemistry*, **37**, 4767–4772.

Griko,Y.V., Friere,E., Privalov,G., van Dael,H. and Privalov,P.L. (1995) *J. Mol. Biol.*, **252**, 447–459.

Grobler J.A., Rao,K.R., Pervaiz,S. and Brew,K. (1994a) *Arch. Biochem. Biophys.*, **313**, 360–366.

Grobler,J.A., Wang,M., Pike,A.C. and Brew,K. (1994b) *J. Biol. Chem.*, **269**, 5106–5114.

Haezebrouck,P., Noppe,W., Van Dael,H. and Hanssens,I. (1992) *Biochem. Biophys. Acta*, **1122**, 305–310.

Harata,K. and Muraki,M. (1992) *J. Biol. Chem.*, **267**, 1419–1421.

Hendrix,T.A., Griko,Y. and Privalov,P. (1996) *Protein Sci.*, **5**, 923–931.

Hunenberger,P.H., Mark,A.E. and van Gunsteren,W.F. (1995) *J. Mol. Biol.*, **252**, 492–503.

Jernigan,R., Raghunathan,G. and Bahar,I. (1994) *Curr. Opin. Struct. Biol.*, **4**, 256–263.

Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.
Karplus,M. and Post,C.B. (1996) *EXS*, **75**, 111–141.
Khatra,B.S., Herries,D.G. and Brew,K. (1974) *Eur. J. Biochem.*, **44**, 537–560.
Koga,K. and Berliner,L.J. (1985) *Biochemistry*, **24**, 7257–7262.
Kraulis,P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.
Kuhlman,B., Boice,J.A., Wu,W.-J., Fairman,R. and Raleigh,D.P. (1997) *Biochemistry*, **36**, 4607–4615.
Laberge,M., Wright,W.W., Sudhakar,K., Liebman,P.A. and Vanderkooi,J.M. (1997) *Biochemistry*, **36**, 5363–5371.
Lindahl,L. and Vogel,H.J. (1984) *Anal. Biochem.*, **140**, 394–402.
Malinovskii,V.A., Tian,J., Grobler,J.A. and Brew,K. (1996) *Biochemistry*, **35**, 9710–9715.
Maurer,P., Hohenester,E. and Engel,J. (1997) *Curr. Opin. Cell. Biol.*, **8**, 609–617.
McCammon,J.A. and Harvey,S.C. (1987) *Dynamics of Proteins and Nucleic acids*. Cambridge University Press, London.
Mckenzie,H.A. (1996) *EXS*, **75**, 365–409.
Morozova,L., Haezebrouk,P. and van Cauwelaert,F. (1991) *Biophys. Chem.*, **41**, 185–191.
Morozova-Roche,L.A., Arico-Muendel,C.C., Haynie,D.T., Emelyanenko,V.I., Dael,H.V. and Dobson,C.M. (1997) *J. Mol. Biol.*, **268**, 903–921.
Nelson,M.R. and Chazin,W.J. (1998) *Protein Sci.*, **7**, 270–282.
Pike,A.C., Brew,K. and Acharya,K.R. (1996) *Structure*, **4**, 691–703.
Post,C.B., Brooks,B.R., Karplus,M., Dobson,C.M., ArtymiukP.J., Cheetham,J.C. and Philips,D.C. (1986) *J. Mol. Biol.*, **190**, 455–479.
Post,C.B., Dobson,C.M. and Karplus,M. (1989) *Protein Struct. Funct. Genet.*, **5**, 337–354.
Qasba,P.K. and Kumar,S. (1997) *Crit. Rev. Biochem. Mol. Biol.*, **32**, 255–306.
Ren,J., Stuart,D.I. and Acharya,K.R. (1993) *J. Biol. Chem.*, **268**, 19292–19298.
Rodriguez,R., Menendez-Arias,L., Gonzalez de Buitrago,G. and Gavilanes,J.,G. (1985) *Biochem. Int.*, **11**, 841–843.
Schulman,B.A., Redfield,C., Peng Z-.Y., Dobson,C.M. and Kim,P.S. (1995) *J. Mol. Biol.*, **253**, 651–657.
Smith,S., Barber,K.R., Dunn,S.D. and Shaw,G.S. (1996) *Biochemistry*, **35**, 8805–8814.
Smith,L.J., Mark,A.E., Dobson,C.M. and van Gunsteren,W.F. (1995) *Biochemistry*, **34**, 10918–10931.
Song,H., Inaka,K., Maenaka,K. and Matsushima,M. (1994) *J. Mol. Biol.*, **244**, 522–540.
Spyracopoulos,L., Li,M.X., Sia,S.K., Gange,S.M., Chandra,M., Solaro,R.J. and Sykes,B.D. (1997) *Biochemistry*, **36**, 12138–12146.
Strynadka,N.C. and James,M.N. (1991) *J. Mol. Biol.*, **220**, 401–424.
Strynadka,N.C. and James,M.N. (1996) *EXS*, **75**, 185–222.
Stuart,D.I., Acharya,K.R., Walker,N.P., Smith,S.G., Lewis,M. and Phillips,D.C. (1986) *Nature*, **324**, 84–87.
Sugai,S. and Ikeguchi,M. (1994) *Adv. Biophys.*, **30**, 37–84.
Tsuge,H., Koseki,K., Miayamo,M., Shimazaki,K., Chuman,T., Matsumoto,T., Noma,M., Nitta,K. and Sugai,S. (1991) *Biochem. Biophys. Acta.*, **1078**, 77–84.
Tsuge,H., Ago,H., Noma,M., Nitta,K., Sugai,S. and Miyano,M. (1992) *J. Biol. Chem. (Tokyo)*, **111**, 141–143.
Teahan,C.G., McKenzie,H.A., Shaw,D.C. and Griffiths,M. (1991) *Biochem. Int.*, **24**, 85–95.
Vanderheeren,G., Hanssens,I., Meijberg,W. and van Aerschot,V. (1996) *Biochemistry*, **35**, 16753–16759.
Wu,L.C., Schulman,B.A., Peng,Z.-Y. and Kim,P.S. (1996) *Biochemistry*, **35**, 859–863.
Zavodszky,P., Kardos,J., Svingor,A. and PetskoG.A. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 7406–7411.

Received January 5, 1998; revised October 14, 1998; accepted October 19, 1998